Fixed effects, random effects, repeated measures

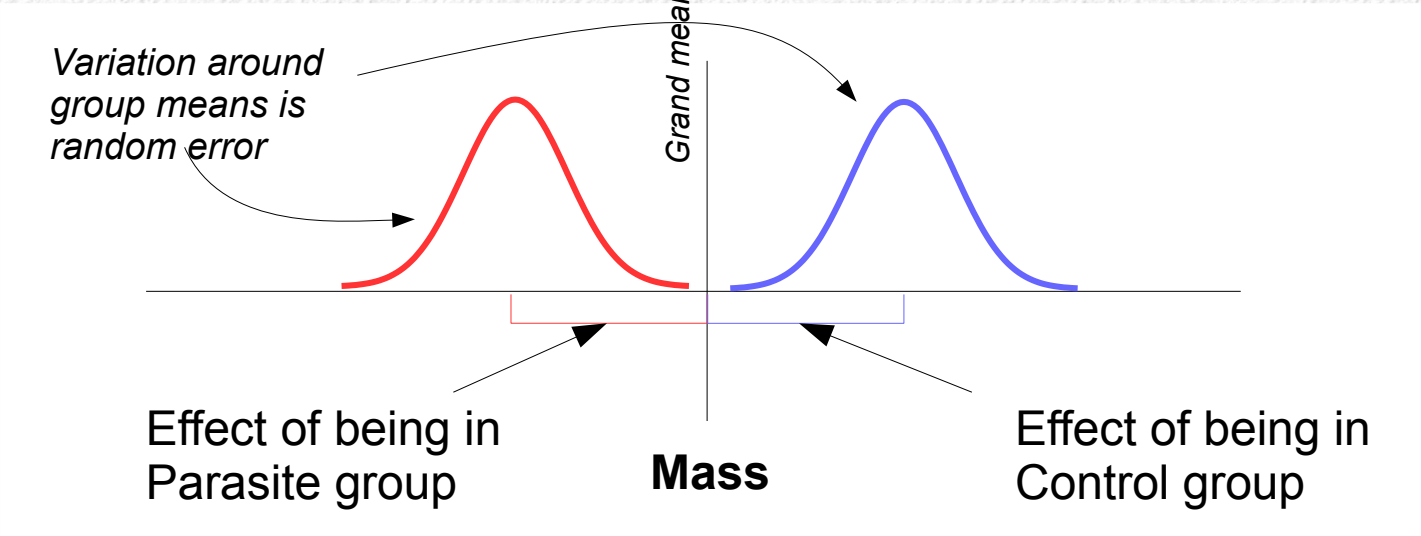
Fixed effects GLM

- So far, everything we have done is a fixed effects GLM (Model 1)
  - Fixed effects = deterministic, non-random effects of a predictor
  - Can be differences between treatment groups (ANOVA), or numeric relationship between predictor and response (regression)
  - Model coefficients are estimates of fixed effects, which are tested against 0
- Random variation interferes with our measurement of the deterministic effect
- Thus, we estimate the size of a fixed effect, imperfectly, subject to random sampling error

Example: effects of intestinal parasites on mass of mice

- Question: does parasite treatment affect average mouse weight?
- Experiment:
  - 20 mice randomly assigned to two treatment groups:
    - Parasite larvae (n = 10)
    - Control (n = 10)
  - Each mouse weighed at the end of 1 week
- An F test of MS_F/MS_E is supposed to tell us whether there is a fixed effect of parasitism
- How, exactly?

Causes of variation in the data



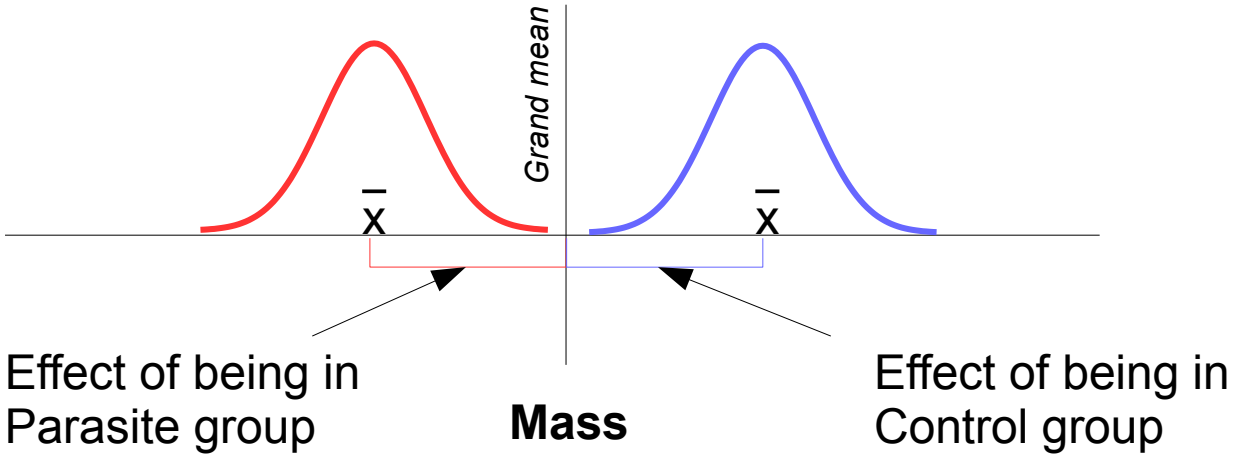
Fixed effects model:

Y = Fixed effect + error

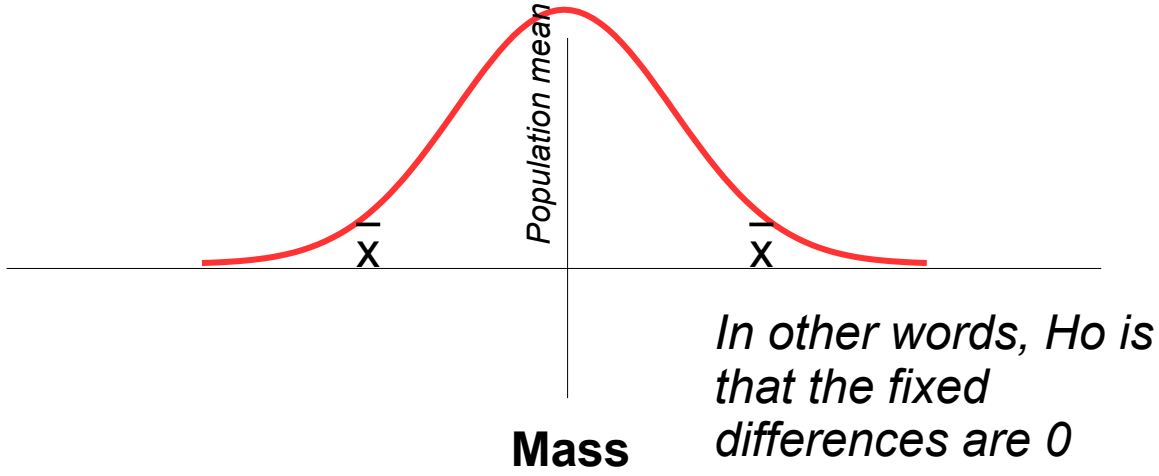
That is, we treat each mouse's mass as a fixed effect of the parasite group the mouse belongs to plus an unpredictable, random individual contribution

What we test with a fixed effects model

Is it this (H_A)?



Or this (H_o)?



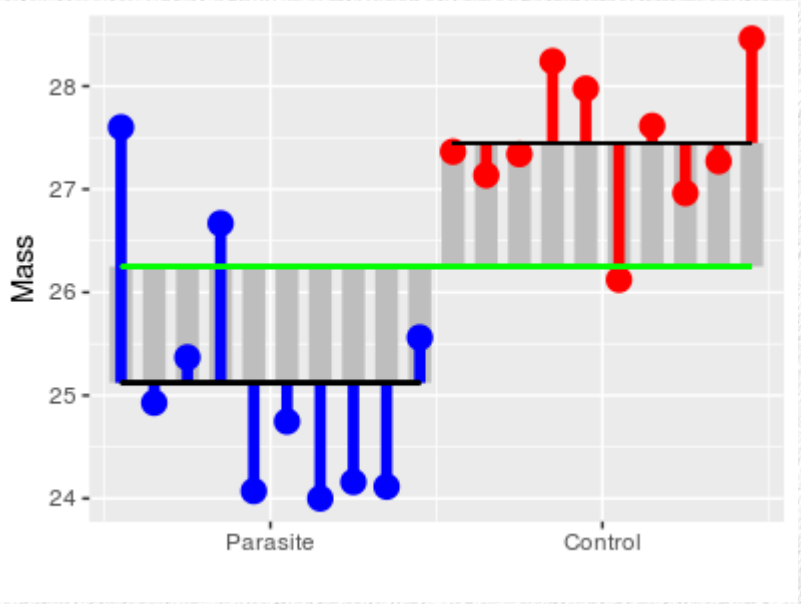
How does MS_F / MS_E tell us if there are fixed effects of treatment?

Expected mean squares

- **Expected mean squares (E(MS))** = the sources of variation that contribute to a MS calculation
  - Include any fixed effects
  - Include any source of randomness that contributes variation to the estimate of the MS for the term

Expected mean squares for the fixed effect of parasite treatment

- Effect of treatment is assessed based on group means
- Variation between the group means is due to:
  - Fixed effects of treatment (γ^2)
  - Individual random variation around group means (σ^2)
- Expressed as an *expected* MS for the fixed effect



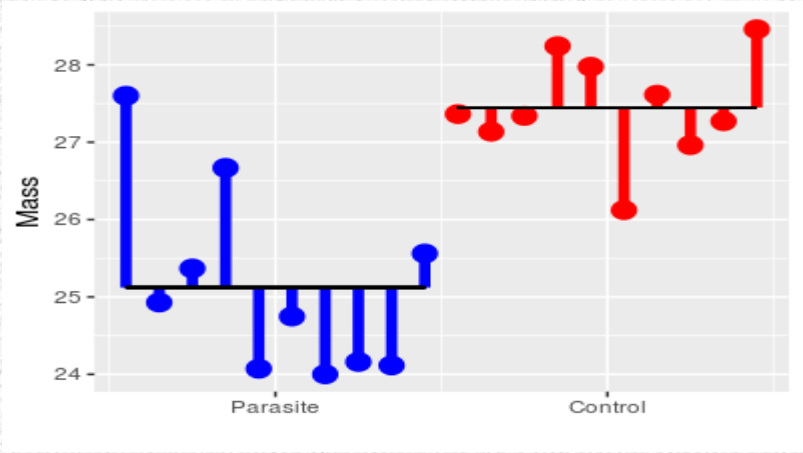
$$E(MS)_{\text{fixed effect}} = \frac{n\sum \gamma^2}{n-1} + \sigma_E^2$$

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treatment	1	27.08	27.078	28.05	4.92e-05
Residuals	18	17.38	0.965		

Estimated by...

Expected mean squares for random error

- Variation around group means is all individual, random variation
- Since individual random variation is the only contribution, the expected value is:



	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treatment	1	27.08	27.078	28.05	4.92e-05
Residuals	18	17.38	0.965		

$$E(MS)_{\text{error}} = \sigma_E^2$$

Estimated by...

Why the F test works

- We use the ratio of two variance estimates (MS_F and MS_E) to test for effects of treatment
- Only difference between MS_F and MS_E is that MS_F contains the fixed effects
- What would $F_{\text{fixed effect}}$ be if the gammas are 0?

$$E(MS)_{\text{fixed effect}} = \frac{n\sum \gamma^2}{n-1} + \sigma_E^2$$

$$E(MS)_{\text{error}} = \sigma_E^2$$

$$F_{\text{fixed effect}} = \frac{\frac{n\sum \gamma^2}{n-1} + \sigma_E^2}{\sigma_E^2} = \frac{MS_F}{MS_E}$$

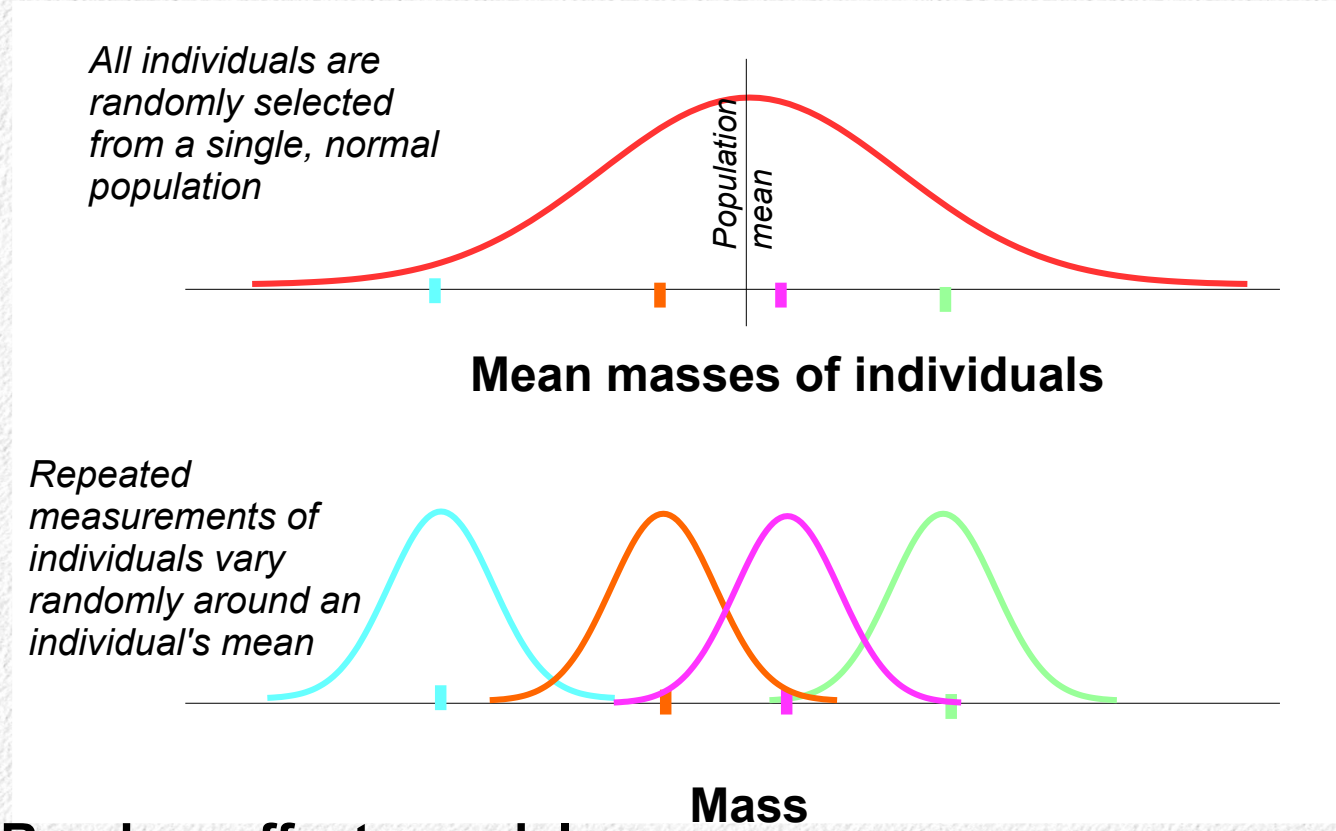
Random effects

- Different mouse experiment:
  - Randomly select 20 mice
  - Measure each one daily for 11 days
- No fixed effects (no treatment), but there are two different levels of random variation
  - Mice differ randomly in size – a **random effect**
  - Each daily measurement of a mouse can be different – error variation
- A random effect is just another identifiable level of random variation, selected from a distribution
- An ANOVA with only random effects is called Model 2 ANOVA
- SS and MS calculations are done the same way as in Model 1 ANOVA, but the interpretation is different

Mouse random effect

Two levels of random variation:

- Among individuals (on average) = mouse random effect
- Among repeated measurements within individuals = error



Random effects model:

Y = Random effect + error

That is, each measurement is due to to the size of the mouse plus random variation in individual daily measurements

EMS for random effects

- All variation is random, so use σ as symbol for both levels
- EMS for the mouse level – affected by the variation among individuals, and the random variation among daily measurements

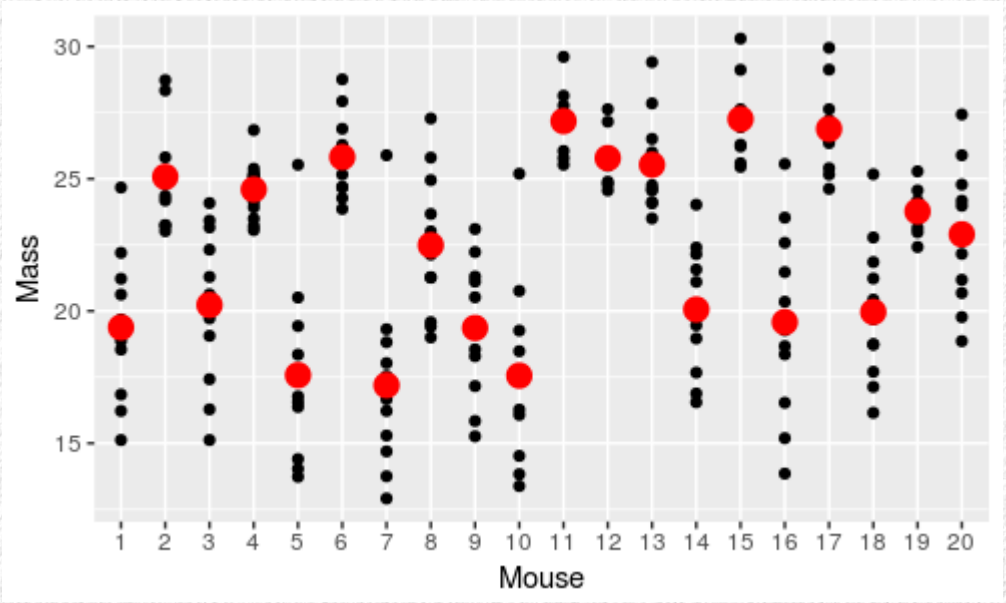
$$E(MS)_{\text{random effect}} = n\sigma_R^2 + \sigma_E^2$$

- EMS for error is still just random variation among daily measurements (i.e. random variation at the lowest level of grouping)

$$E(MS)_{\text{error}} = \sigma_E^2$$

The data

- Mouse level
  - Mean masses for the mice vary → random effect of mouse
- Error level
  - Daily measurements of each mouse vary around the mouse's mean



F test for a random effect

- Random effects F test
  - The expected MS tell us to divide the mouse term by the error term
  - Isolates variation among mice
- But, if we run the ANOVA in R...

$$F_{\text{random effect}} = \frac{n\sigma_R^2 + \sigma_E^2}{\sigma_E^2} = \frac{MS_{mouse}}{MS_E}$$

Random effects ANOVA table

summary(aov(Mass ~ Error(Mouse), data=crmasses))

```
Error: mouse
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 19  2530   133.1

Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 200  1176  5.881
```

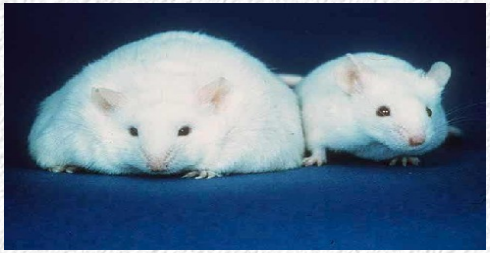
Need to use the aov() command instead of lm()

Error() indicates Mouse is a random effect (an “error strata”, in R lingo)

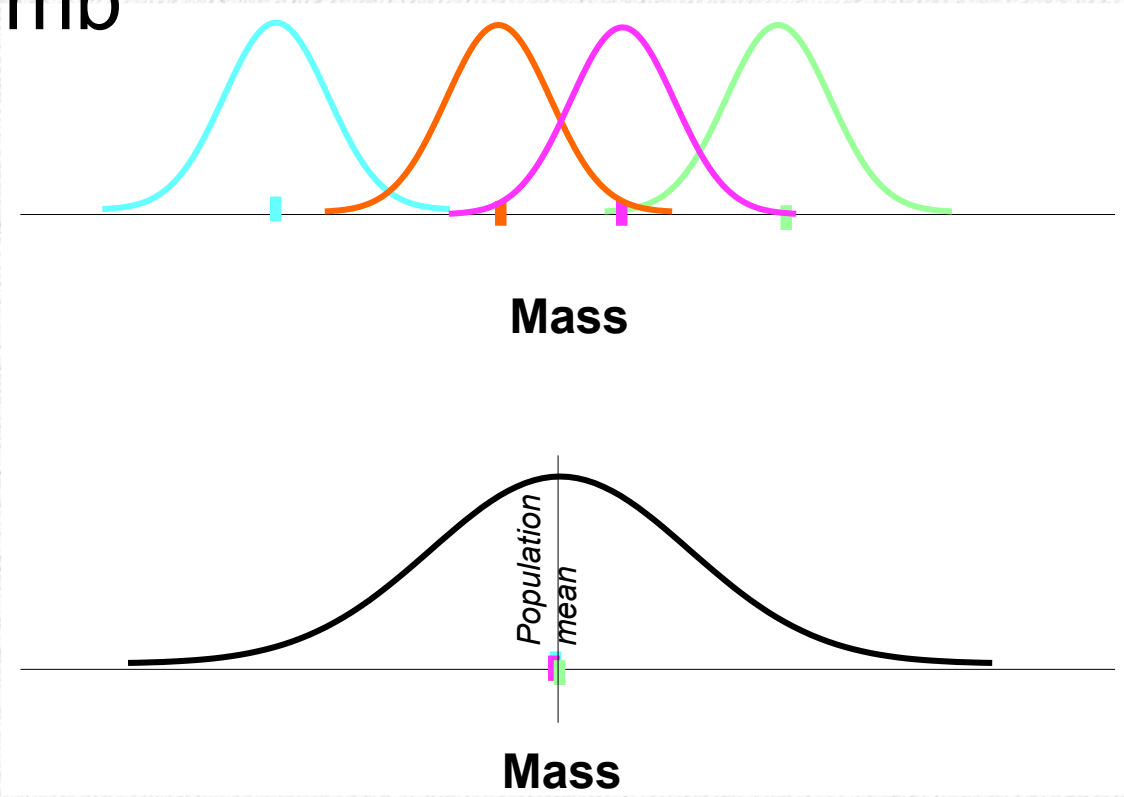
No p-values...why? Because R programmers think p-values for random effects are stupid

Why R thinks p-values on random effects are dumb

Is it this (H_A)?



Or this (H_o)?



What are the chances that the null is actually true? Is there any reason to test hypotheses we know are false?

Preferred analysis for random effects: variance components

- Treat as an *estimation* problem, not as a hypothesis testing problem
- Estimates of the amount of random variation at each level = **variance components**
- Components are σ^2_{mouse} and σ^2_{error}
- R ANOVA table gives us:

  MS_{mouse} = 133.1 → estimate of: $E(MS)_{mouse} = n\sigma^2_{mouse} + \sigma^2_{error}$

  MS_{error} = 5.881 → estimate of: $E(MS)_{error} = \sigma^2_{error}$
- With a little algebra...

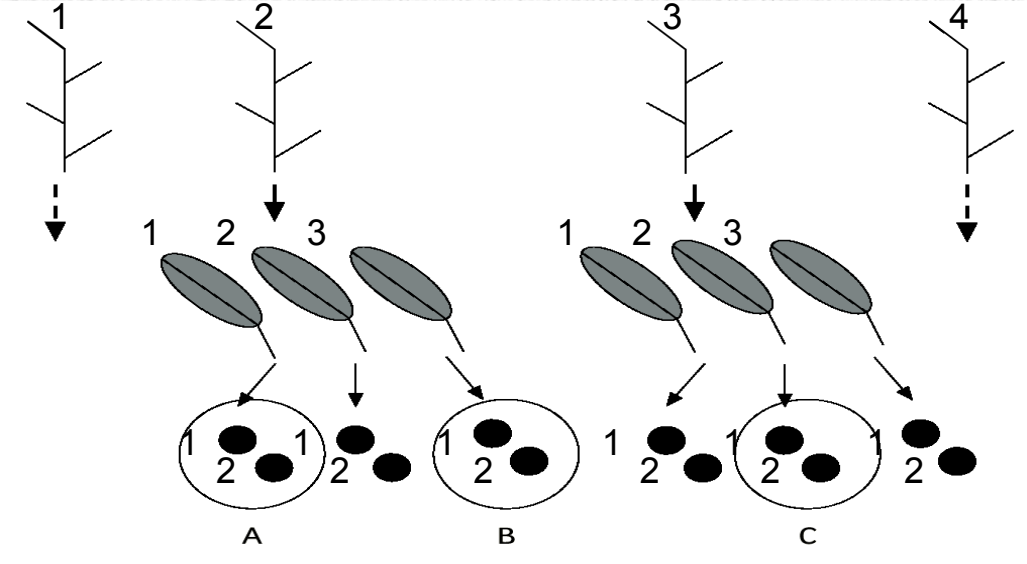
$$\sigma^2_{\text{mouse}} = \frac{MS_{\text{mouse}} - MS_{\text{error}}}{n} = \frac{133.1 - 5.881}{11} = 11.57$$

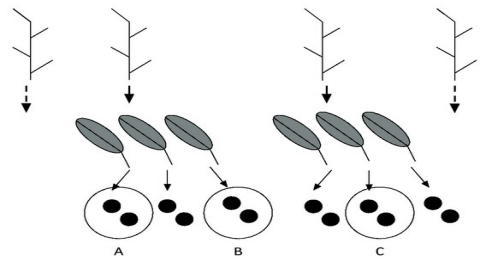
$$\sigma^2_{\text{error}} = 5.881$$

Conclusion: about twice as much random variation among mice as within

Nested designs

- It isn't always possible to cross factors – levels may be nested inside of each other
- Example: measurements of leaf calcium concentration in plants
  - Four plants selected for measurement – levels Plant 1, Plant 2, Plant 3, Plant 4
  - Three leaves from each plant selected – levels Leaf 1, Leaf 2, Leaf 3
  - Two discs cut out of each leaf, measured for calcium concentration
- Leaves are nested within plant
- Discs are nested within leaf
- Random effect of plant, random effect of leaf (disc is error variation)
- Note – numbers are just identifiers (Leaf 1 not a treatment level, just first leaf from each plant)





EMS for each level

Expected MS for:

Plant	$nm\,\sigma^2_{\text{plant}} + n\,\sigma^2_{\text{leaf}} + \sigma^2_E$	*Variation of plant means around the grand mean*
Leaf	$n\,\sigma^2_{\text{leaf}} + \sigma^2_E$	*Variation of leaf means around their plant mean*
Error	σ^2_E	*Variation of discs around their leaf means*

ANOVA table

```
Call:
aov(formula = Conc ~ Error(Plant/Leaf), data = plants)

Error: Plant
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  3  2.724  0.9081

Error: Plant:Leaf
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  8   1.27  0.1587

Error: Within
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 12 0.5083 0.04236
```

Variance components

$$EMS_{\text{plant}} = nm\,\sigma^2_{\text{plant}} + n\,\sigma^2_{\text{leaf}} + \sigma^2_E \qquad MS_{\text{plant}} = 0.9081$$

$$EMS_{\text{leaf}} = n\,\sigma^2_{\text{leaf}} + \sigma^2_E \qquad MS_{\text{leaf}} = 0.1587$$

$$EMS_{\text{error}} = \sigma^2_E \qquad MS_{\text{error}} = 0.04236$$

$$\sigma^2_{\text{plant}} = \frac{MS_{\text{plant}} - MS_{\text{leaf}}}{nm} = \frac{0.9081 - 0.1587}{2 \times 3} = 0.1249$$

Interpretation – greatest variability at plant level, about equal amounts of variability at the leaf and individual disc level

$$\sigma^2_{\text{leaf}} = \frac{MS_{\text{leaf}} - MS_{\text{error}}}{n} = \frac{0.1587 - 0.04236}{2} = 0.0582$$

So? In an experiment, may need lots of plants but not very many leaves per plant, discs per leaf

Mixing fixed and random (Model 3 ANOVA)

- What if we:
  - Used 10 parasitized, 10 control mice
  - Weighed them daily for 11 days
- We have a parasite treatment (fixed) and several measurements for individual mice (random) → a **mixed model** (Model 3 ANOVA)
- Problem: which of the two different levels of random variation (mouse, error) should we use to test the parasite effect?
- We can use EMS to tell us which MS is the right denominator for our test of the fixed effect of parasite

Mixed model EMS

$$E\left(MS_{par.trt}\right)=\frac{n\sum\gamma^2}{n-1}+n\sigma^2_{mouse}+\sigma^2_E$$

Effect of parasite subject to random variation between mice and within mice (error)

$$E\left(MS_{mouse}\right)=n\sigma^2_{mouse}+\sigma^2_E$$

EMS for mouse includes both between-mouse and within-mouse

$$E\left(MS_{error}\right)=\sigma^2_E$$

EMS for repeated measurements of each mouse (error)

What's the appropriate denominator to test the fixed effect of parasite treatment?

Mixed effects model in R

aov(Mass ~ Parasite + Error(Mouse))

- Mouse is specified as a random effect with Error(Mouse)
- R knows to calculate a separate error term for Mouse and Residuals
- Also knows that Parasite should be tested over the Mouse term

```
Error: Mouse
          Df Sum Sq Mean Sq F value   Pr(>F)
Parasite   1 2019.3  2019.3   71.24 1.13e-07
Residuals 18  510.2    28.3

Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 200   1176   5.881
```

Using Mouse as the error term is equivalent to averaging by mouse, then using the averages as the data to test for parasite effects

Parasite effect using means for each mouse

```
            Df Sum Sq Mean Sq F value   Pr(>F)
Parasite     1 183.57  183.57   71.24 1.13e-07
Residuals   18  46.39    2.58
```

Variance components for mouse and residual

```
Error: Mouse
          Df Sum Sq Mean Sq F value   Pr(>F)
Parasite   1 2019.3  2019.3   71.24 1.13e-07
Residuals 18  510.2    28.3

Error: Within
           Df Sum Sq Mean Sq F value Pr(>F)
Residuals 200   1176   5.881
```

$$\sigma^2_{\text{Mouse}} = \frac{MS_{\text{Mouse}} - MS_{\text{residual}}}{n} = \frac{28.3 - 5.881}{11} = 2.04$$

So, 2.04/5.88 = 0.35, so random variation between mice is 35% the size of random variation between repeated measurements of the same mouse

Problem: repeated measures

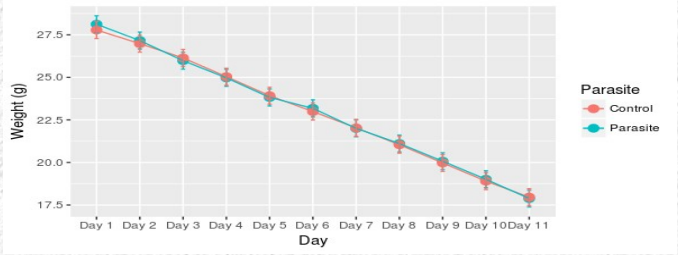
- Repeated measurements of individuals are not independent
  - Pseudoreplication if we use each measurement as a replicate
  - May be patterning in residuals due to change over time
- Using mouse as a random effect only solves the pseudoreplication problem
- We may also be experimentally interested in the change over time
  - When do the treatments become different from one another?
  - How do the patterns of change compare between the treatments?
- Accounting for serially dependent measurements is done with **repeated measures ANOVA**

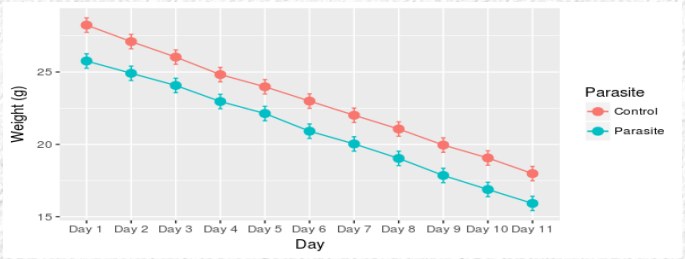
The repeated measures design

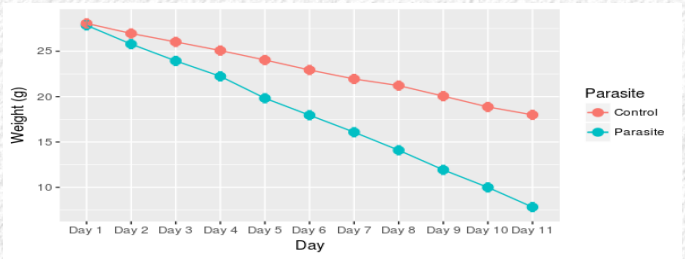
- Distinguishes **between-subjects** and **within-subjects** effects
  - Between subjects = treatments applied to different subjects (each subject can only be in one group)
  - Within subjects = measurements taken on every subject for every level (i.e. each subject measured every day)
- Parasite treatment is the between-subjects treatment
- Time is the within-subjects treatment

Main effects and interactions in RMA

- Main effect of parasite indicates that masses differ between parasite and control groups
- Main effect of time indicates change in average mass over time
- Interaction of parasite by time (between x within) indicates that the change over time depends on treatment







Organization of data

- Each row is an observation of an individual
- Column for Time indicating the time of the measurement (as a factor)
- Mouse column used as a random effect (as a factor)

Mouse	Parasite	Mass	Time
15	Control	30.30	1
8	Control	27.28	1
17	Control	29.95	1
13	Control	29.41	1
12	Control	27.63	1
11	Control	29.61	1
4	Control	26.84	1
19	Control	25.28	1
2	Control	28.34	1
6	Control	28.76	1
20	Treatment	24.17	1
18	Treatment	25.17	1
16	Treatment	25.56	1
14	Treatment	22.15	1
5	Treatment	25.53	1
3	Treatment	23.16	1
1	Treatment	24.67	1
10	Treatment	25.19	1
9	Treatment	21.11	1
7	Treatment	25.89	1
15	Control	29.12	2
8	Control	25.80	2
17	Control	29.13	2

Repeated measures as a mixed effects model in R

aov(Mass ~ Parasite*Time + Error(Mouse))

```
Error: Mouse
          Df Sum Sq Mean Sq F value   Pr(>F)
Parasite   1 2019.3  2019.3   71.24 1.13e-07 ***
Residuals 18  510.2    28.3

Error: Within
              Df Sum Sq Mean Sq F value Pr(>F)
Time          10  937.9   93.79  129.29 <2e-16 ***
Parasite:Time 10  107.8   10.78   14.86 <2e-16 ***
Residuals    180  130.6    0.73
```

Each mouse measured at each time point, so time only subject to error variation

An additional assumption of RMA: sphericity

- Sphericity = variances between successive time points are the same
- Sphericity needed for the p-values for the time and time x parasite interactions to be accurate
- If we violate sphericity, we need to:
  - Adjust the p-values (if we don't violate it too badly)
  - Use an approach that doesn't require sphericity called “profile analysis” (if we violate it badly)

Testing for sphericity, correcting for lack of it

- Tested with the Mauchly test (if $p < 0.05$, fail the test)
- Three common corrections – Greenhouse-Geisser, Huynh-Feldt, and lower-bound
  - All three based on “epsilon”, which is 1 when variances are identical across time points, approaches 0 as the variance become increasingly different
  - Epsilon > 0.9 → sphericity is met, p-values in ANOVA table accurate
  - 0.9 > epsilon > 0.7 → sphericity is violated, but can use adjusted p-values
  - Epsilon < 0.7 → use profile analysis (multivariate approach)… less powerful than univariate repeated measures when assumptions met, but better option when assumptions not met

Sphericity failed, use corrected p-values, or profile analysis

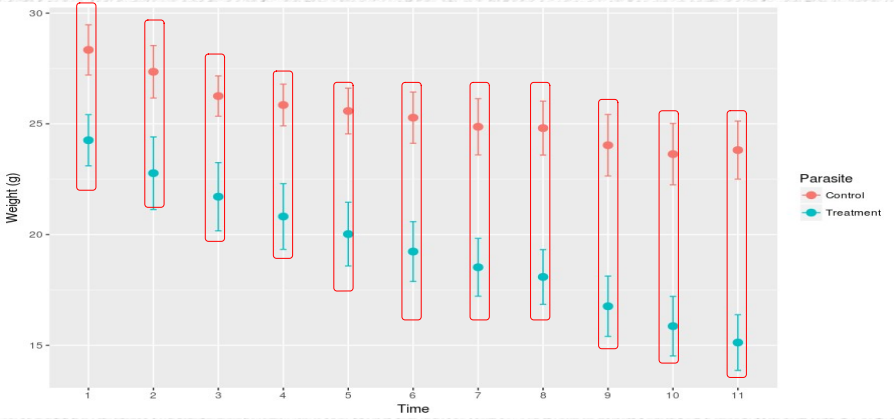
```
Mauchly Tests for Sphericity
                           Test statistic   p-value
time.factor                    4.8084e-09 1.033e-30
masses$Parasite:time.factor    4.8084e-09 1.033e-30



Greenhouse-Geisser and Huynh-Feldt Corrections
 for Departure from Sphericity
                            GG eps Pr(>F[GG])
time.factor                0.21105  < 2.2e-16 ***
masses$Parasite:time.factor 0.21105  1.265e-05 ***

                            HF eps Pr(>F[HF])
time.factor                0.23978  < 2.2e-16 ***
masses$Parasite:time.factor 0.23978  3.986e-06 ***
```

Post-hocs in repeated measures: with no between x within interaction

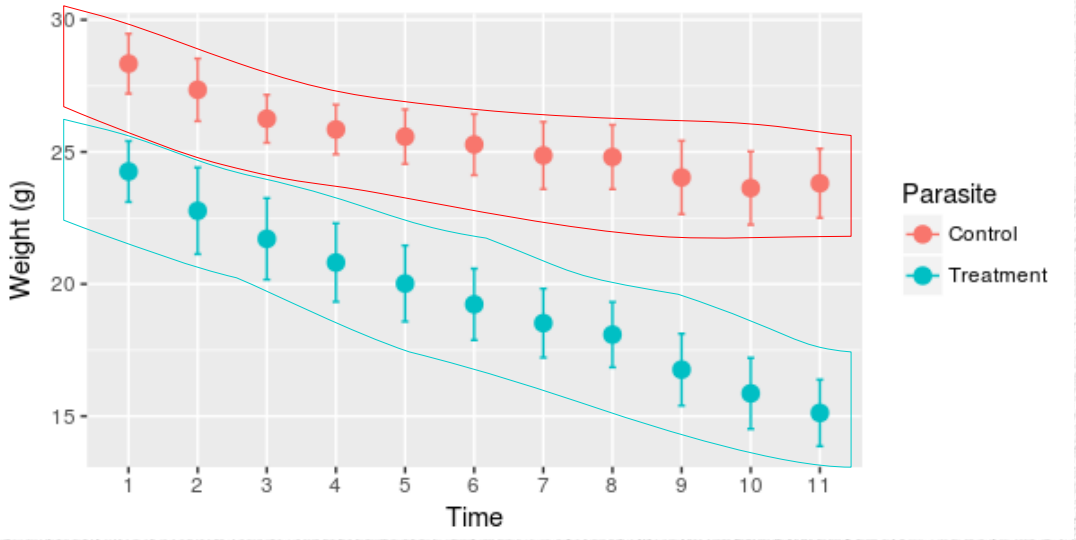
- Within-subjects (time) = compare time points
  - Repeated measurements are paired by individual, should base on **paired** t-tests (why?)
    - Can be all possible pairs of time points
    - Can be only sequential differences
    - Can be initial conditions vs. each subsequent
  - Can use orthogonal polynomials to assess trends over time
- Between-subjects (parasite) – test with Tukey tests

Post-hocs with an interaction

- Comparing all possible time points between treatments *not* usually desirable
  - 22 combinations of treatment x time
  - 231 pairs of means
- Can assess differences in time trend with orthogonal polynomials
- Can test differences in treatment means at each time point – find the time points at which treatment groups differ



- Can test differences between time points within each treatment group
  - Paired analysis, data subset by parasite group
  - All possible, sequential, against initial



In either case, would want to use an adjusted alpha level to account for the number of comparisons – α/k